

NO-A189 053

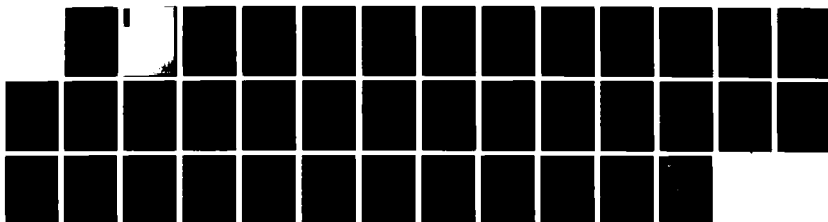
SURVIVAL ANALYSIS USING ADDITIVE RISK MODELS(U)
STANFORD UNIV CA DEPT OF STATISTICS F W HUFFER ET AL.
06 OCT 87 TR-396 N00014-86-K-0156

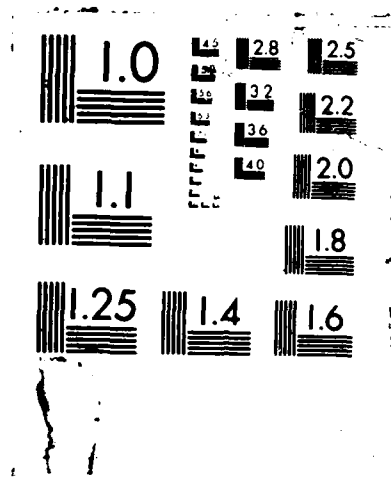
1/1

UNCLASSIFIED

F/G 12/3

NL





3
3
3
3
3

SURVIVAL ANALYSIS USING ADDITIVE RISK MODELS

BY

FRED W. HUFFER and IAN W. McKEAGUE

TECHNICAL REPORT NO. 396

OCTOBER 6, 1987

Prepared Under Contract

N00014-86-K-0156 (NR-042-267)

For the Office of Naval Research

Herbert Solomon, Project Director

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Accession For	
NTIS CRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

1. INTRODUCTION

→ In this paper we study Aalen's (1980) additive risk model for the regression analysis of censored survival data. Let $\lambda_i(t) = \lambda_i(t, Y_i)$ denote the hazard function at time t for subject i whose covariates are given by the p -vector $Y_i = (Y_{i1}, \dots, Y_{ip})'$. Aalen's model stipulates that

$$\lambda_i(t) = \sum_{j=1}^p \alpha_j(t) Y_{ij} = Y_i' \alpha(t) \quad (1.1)$$

where $\alpha = (\alpha_1, \dots, \alpha_p)'$ is an unknown vector of hazard functions. More generally, the covariates can be time dependent, as considered in section 5.

→ The additive risk model provides a useful alternative to Cox's (1972) proportional hazards model when large sample size makes its application feasible. It is capable of providing detailed information concerning the temporal influence of each covariate. The temporal influences of the covariates are not assumed to be proportional as they are in Cox's model. Buckley (1984) has pointed out that additive risk models are biologically more plausible than proportional hazard models. Also, the use of the proportional hazards model when the true model is additive risk has been found by O'Neill (1986) to result in serious asymptotic bias.

Aalen (1980) introduced estimators for the vector of integrated hazard functions, $A(t) = \int_0^t \alpha(s) ds$, which use continuous data (containing the exact values of failure and censoring times). These estimators generalize the well-known Nelson-Aalen estimator, the natural estimator in the case of one covariate. However, except in the case of one covariate (Aalen, 1978), the asymptotic theory was not fully developed. One possible form of these estimators was motivated by a formal least squares principle. This estimator, defined precisely in (2.3) and (5.4), is referred to here as *Aalen's least squares estimator*. Aalen observed that this estimator probably gives reasonable estimates and he applied it to analysis of data from the Veterans Administration Lung Cancer Study Group.

→ The first purpose of the present paper is to apply the additive risk model to the analysis of *grouped data* in which only the person-years at risk and number of uncensored

deaths over successive time intervals, tabulated for various levels of the covariates, are available. This kind of data typically arises in epidemiological cohort studies involving the follow-up of large population groups over many years, see Breslow (1986). Our approach is to use an estimator, constructed using the method of sieves (Grenander, 1981), for which an asymptotic distribution theory was developed by McKeague (1987). This estimator, called the integrated histogram sieve estimator, requires only grouped data.

In section 2 we describe the various estimators and confidence bands, give a heuristic motivation for the integrated histogram sieve estimator and show that it is approximately unbiased. The results of a simulation study are reported in section 3. In section 4 we apply the additive risk model to the analysis of grouped data on the incidence of cancer mortality among Japanese atomic bomb survivors.

The second purpose of this paper is to derive the asymptotic distribution of Aalen's least squares estimator. This is done in Section 5. It turns out that, appropriately normalized, Aalen's least squares estimator and the integrated histogram sieve estimator have the same asymptotic distribution. In comparing the two estimators this indicates that (asymptotically) there is no loss in using the grouped data when the continuous data is unavailable. We conclude section 5 with a discussion of *weighted* least squares estimators for the additive risk model.

2. ESTIMATORS AND CONFIDENCE BANDS

We shall use the random censorship model in which the i th individual's failure time X_i is assumed to be an absolutely continuous random variable conditionally independent of the censoring time W_i given the covariate vector Y_i . Let $\tilde{X}_i = \min(X_i, W_i)$ and $\delta_i = I(X_i \leq W_i)$ denote the time to the end-point event and the indicator for noncensorship respectively. Assume that the observation triples $(\tilde{X}_i, \delta_i, Y_i), i = 1, \dots, n$ are i.i.d. and the conditional hazard function $\lambda_i(t)$ of X_i given Y_i satisfies the model (1.1).

Let I_1, \dots, I_d be intervals which partition the follow-up period $[0, T]$ over which estimation of $\alpha_1, \dots, \alpha_p$ or A_1, \dots, A_p is desired. Specifically, let $I_r = (b_r, b_r + \ell_r]$,

$r = 1, \dots, d$. The total time that the i th individual is observed to be at risk in interval I_r is given by

$$T_{ir} = \int_{I_r} I(\tilde{X}_i \geq t) dt.$$

The indicator that the i th individual undergoes an uncensored failure in I_r is given by

$$\delta_{ir} = \delta_i I(\tilde{X}_i \in I_r).$$

Let C_r be the $p \times 1$ vector

$$C_r = \sum_{i=1}^n \delta_{ir} Y_i$$

and let D_r be the $p \times p$ matrix

$$D_r = \sum_{i=1}^n (Y_i Y_i') T_{ir}.$$

In the sequel we use the following notational convention: for any square matrix K , K^{-1} denotes the inverse of K if K is invertible, the zero matrix otherwise. The histogram sieve estimator is defined by $\hat{\alpha}(t) = \hat{\alpha}_r$ when $t \in I_r$, where $\hat{\alpha}_r = (\hat{\alpha}_{rj})$ is the $p \times 1$ vector given by $\hat{\alpha}_r = D_r^{-1} C_r$. Note that the histogram sieve estimator can be evaluated from the grouped data consisting of the total time at risk and the number of uncensored failures in each interval I_1, \dots, I_d tabulated for all realized levels of the covariates.

The form of the histogram sieve estimator can be motivated by the following argument. Suppose the time intervals I_1, I_2, \dots, I_d are short enough so that the hazard rates $\lambda_i(t)$ for all subjects can be regarded as constant within each interval $I_r = (b_r, b_r + \ell_r]$, that is, we can take $\alpha(t) = \alpha_r$ for $t \in I_r$. This implies $\lambda_i(t) = \lambda_{ir}$ for $t \in I_r$ where $\lambda_{ir} = Y_i' \alpha_r$.

Under this condition we can show that $E(D_r \alpha_r | \mathcal{F}_r) = E(C_r | \mathcal{F}_r)$ where \mathcal{F}_r is the σ -field generated by the covariates Y_1, \dots, Y_n and the events $\{\tilde{X}_i > b_r\}$, $i = 1, \dots, n$. If the number of observed failures in I_r is large, the Law of Large Numbers will imply $D_r \alpha_r \approx C_r$. This suggests the estimator $\hat{\alpha}_r = D_r^{-1} C_r$ which is the histogram sieve estimator.

To show that $E(D_r \alpha_r | \mathcal{F}_r) = E(C_r | \mathcal{F}_r)$ we shall prove the stronger result $E(D_r \alpha_r | \mathcal{G}_r) = E(C_r | \mathcal{G}_r)$ where \mathcal{G}_r is the larger σ -field generated by \mathcal{F}_r and the censoring times

W_1, \dots, W_n . From the definitions we have

$$E(C_r | \mathcal{G}_r) = \sum_{i=1}^n Y_i E(\delta_{ir} | \mathcal{G}_r) \quad \text{and}$$

$$E(D_r \alpha_r | \mathcal{G}_r) = \sum_{i=1}^n Y_i (Y_i' \alpha_r) E(T_{ir} | \mathcal{G}_r).$$

Let $Z_{ir} = (W_i - b_r) \wedge \ell_r$. Given $\tilde{X}_i > b_r$, the time T_{ir} is an exponential random variable (with parameter λ_{ir}) truncated at Z_{ir} so that we obtain

$$E(T_{ir} | \mathcal{G}_r) = \lambda_{ir}^{-1} [1 - \exp(-\lambda_{ir} Z_{ir})] I(\tilde{X}_i > b_r).$$

Similarly

$$E(\delta_{ir} | \mathcal{G}_r) = P\{\tilde{X}_i > b_r, T_{ir} < Z_{ir} | \mathcal{G}_r\}$$

$$= [1 - \exp(-\lambda_{ir} Z_{ir})] I(\tilde{X}_i > b_r).$$

Substituting these in the above formulas for $E(C_r | \mathcal{G}_r)$ and $E(D_r \alpha_r | \mathcal{G}_r)$ and noting that $Y_i' \alpha_r = \lambda_{ir}$, we obtain the desired equality.

Let m_r be the number of subjects at risk in I_r , this is $m_r = \sum_i I(\tilde{X}_i > b_r)$. The histogram sieve estimator $\hat{\alpha}_r$ will be approximately unbiased as long as m_r is large and $\lambda_{ir} \ell_r$ is small for all i ; the probability (conditional on $\tilde{X}_i > b_r$) of failure during I_r is small for all subjects i . In showing this we use the notation

$$F_r = E(D_r | \mathcal{G}_r), \quad \psi_r = D_r - F_r \quad \text{and} \quad \Lambda_r = C_r - F_r \alpha_r.$$

Both $E(\psi_r | \mathcal{G}_r) = 0$ and $E(\Lambda_r | \mathcal{G}_r) = 0$ with the second fact following from

$$F_r \alpha_r = E(D_r \alpha_r | \mathcal{G}_r) = E(C_r | \mathcal{G}_r)$$

which was demonstrated previously. Explicit formulas for ψ_r and Λ_r are

$$\psi_r = \sum_{i=1}^n Y_i Y_i' T_{ir}^* \quad \text{and} \quad \Lambda_r = \sum_{i=1}^n Y_i \delta_{ir}^*$$

with $T_{ir}^* = T_{ir} - E(T_{ir} | \mathcal{G}_r)$ and $\delta_{ir}^* = \delta_{ir} - E(\delta_{ir} | \mathcal{G}_r)$.

Since $\lambda_{i,r}\ell_r$ is small, it is easy to see that $T_{i,r}^*$ is small with high probability. This implies ψ_r is small relative to F_r and justifies the series expansion

$$D_r^{-1} = (F_r + \psi_r)^{-1} = F_r^{-1} - F_r^{-1}\psi_r F_r^{-1} + F_r^{-1}\psi_r F_r^{-1}\psi_r F_r^{-1} - \dots$$

Using this expansion and $C_r = F_r\alpha_r + \Lambda_r$ we find (ignoring higher order terms)

$$E(D_r^{-1}C_r|\mathcal{G}_r) \approx \alpha_r + E(F_r^{-1}\psi_r F_r^{-1}\psi_r\alpha_r|\mathcal{G}_r) - E(F_r^{-1}\psi_r F_r^{-1}\Lambda_r|\mathcal{G}_r).$$

Both of the bias terms are $O(m_r^{-1})$. For example, writing ψ_r and Λ_r as sums and using the independence of the subjects gives

$$E(F_r^{-1}\psi_r F_r^{-1}\Lambda_r|\mathcal{G}_r) = \sum_{i=1}^n F_r^{-1}Y_i Y_i' F_r^{-1}Y_i E(T_{i,r}^* \delta_{i,r}^*|\mathcal{G}_r).$$

This sum contains m_r nonzero terms. We expect the norm $\|F_r^{-1}\|$ to be $O(m_r^{-1})$ so that each nonzero term in the sum will be $O(m_r^{-2})$. Thus the sum is $O(m_r \cdot m_r^{-2}) = O(m_r^{-1})$. The other bias term is handled similarly.

Note, we are assuming above that the hazard rates for all subjects are constant within each interval I_r . If the hazard rates vary within the intervals, this will introduce an additional source of bias.

The integrated histogram sieve estimator is denoted $\hat{A}(t) = \int_0^t \hat{\alpha}(s)ds$, where integration of $\hat{\alpha}$ is carried out componentwise. Let $A(t) = \int_0^t \alpha(s)ds$. Under conditions (C1)-(C4) in section 5.2, given that the partition I_1, \dots, I_d becomes finer as $n \rightarrow \infty$ in such a way that $\sqrt{n} \max(\ell_1, \dots, \ell_d) \rightarrow 0$ and $n \min(\ell_1, \dots, \ell_d) \rightarrow \infty$, then an asymptotic $100(1-\alpha)\%$ confidence band for A_j is given by

$$\hat{A}(t) \pm c_\alpha n^{-1/2} \hat{G}_j(T)^{1/2} (1 + \frac{\hat{G}_j(t)}{\hat{G}_j(T)}), t \in [0, T] \quad (2.1)$$

where c_α is the upper α quantile of the distribution of $\sup_{t \in [0, \frac{1}{2}]} |B^0(t)|$, B^0 is the Brownian bridge process,

$$\hat{G}_j(t) = \int_0^t \hat{g}_j(s)ds \quad (2.2)$$

where $\hat{g}_j(t) = \hat{g}_{rj}$ if $t \in I_r$ and

$$\hat{g}_{rj} = n \sum_{u=1}^p \hat{\alpha}_{ru} \sum_{v=1}^p \sum_{w=1}^p \hat{M}_r(u, v, w) (D_r^{-1})_{jv} (D_r^{-1})_{jw},$$

$$\hat{M}_r(u, v, w) = \sum_{i=1}^n Y_{iu} Y_{iv} Y_{iw} T_{ir}.$$

A table for the distribution of $\sup_{t \in [0, \frac{1}{2}]} |B^0(t)|$ has been given by Hall and Wellner (1980).

An asymptotic $100(1 - \alpha)\%$ confidence interval for $A_j(t)$, at fixed $t \in [0, T]$, is given by

$$\hat{A}_j(t) \pm z_{\alpha/2} n^{-1/2} \hat{G}_j(t)^{1/2},$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile for the standard normal distribution.

Aalen's least squares estimator is defined by

$$\tilde{A}(t) = \sum_{\substack{\delta_i=1 \\ \tilde{X}_i \leq t}} \tilde{D}_i^{-1} Y_i \quad (2.3)$$

where the summation is over individuals i whose failure times are uncensored and less than or equal to t , \tilde{D}_i is the $p \times p$ matrix

$$\tilde{D}_i = \sum_{\tilde{X}_k \geq \tilde{X}_i} Y_k Y_k' \quad (2.4)$$

where the summation is over individuals k at risk at time \tilde{X}_i .

Under conditions (A1)-(A3) in section 5.1 an asymptotic $100(1 - \alpha)\%$ confidence band for A_j is given by

$$\tilde{A}_j(t) \pm c_\alpha n^{-1/2} \tilde{G}_j(T)^{1/2} \left(1 + \frac{\tilde{G}_j(t)}{\tilde{G}_j(T)} \right), t \in [0, T] \quad (2.5)$$

where c_α is defined above and

$$\tilde{G}_j(t) = \sum_{\substack{\delta_i=1 \\ \tilde{X}_i \leq t}} \tilde{g}_{ij}, \quad (2.6)$$

$$\tilde{g}_{ij} = n \sum_{u=1}^p (\tilde{D}_i^{-1} Y_i)_u \sum_{v=1}^p \sum_{w=1}^p \tilde{M}_i(u, v, w) (\tilde{D}_i^{-1})_{jv} (\tilde{D}_i^{-1})_{jw},$$

$$\tilde{M}_i(u, v, w) = \sum_{\tilde{X}_k \geq \tilde{X}_i} Y_{ku} Y_{kv} Y_{kw}.$$

An asymptotic $100(1 - \alpha)\%$ confidence interval for $A_j(t)$, at fixed $t \in [0, T]$, is given by $\tilde{A}_j(t) \pm z_{\alpha/2} n^{-1/2} \tilde{G}_j(t)^{1/2}$.

Note that, unlike the confidence intervals, the confidence bands given above are dependent on the choice of T . As T increases the band widens at all points. It also becomes more unreliable since the effective sample size $n_T = \#\{i : \tilde{X}_i \geq T\}$ for estimating the asymptotic variance at time T is getting smaller. In practice we found it reasonable to set T so that n_T is at least 10% of the sample size. This is the case for the simulated data in section 3 and for the atomic bomb survivor data in section 4.

3. ADDITIVE RISK MODEL SIMULATIONS

In order to evaluate the performance of the confidence intervals and bands proposed for grouped data, a Monte-Carlo experiment was performed to see whether their asymptotic properties take effect under sample sizes, grouping and censoring found in typical applications. The simulation model used $p = 2$ covariates with i.i.d. uniform distributions on the lattice $\{\frac{r}{8}, r = 1, \dots, 8\}$ and corresponding hazard functions $\alpha_1(t) = 1, \alpha_2(t) = t, t \geq 0$. The censoring time was independent of the failure time and exponentially distributed with parameter β , for various values of β . The follow-up period was $[0, 1]$. The sieve intervals I_1, \dots, I_d were of equal length. For various combinations of n , d , and β , 500 samples of grouped data were generated via the random censorship model. For each of these 500 samples we constructed the asymptotic 95% confidence bands (2.1) for the true integrated hazard functions $A_1(t) = t$ and $A_2(t) = \frac{1}{2}t^2$ on the interval $[0, 1]$. The proportion among the 500 bands that contained A_j on the interval $[0, 1]$ was then determined (for $j=1$ and 2). The sample size n was set to 1000 and 4000. The sieve dimension d was set to 8, 16, and 64. The censoring parameter β took the values .3 and 1.5 which amounted to 28% and 68% censoring prior to the end of follow-up, respectively. With $\beta = .3$ ($\beta = 1.5$) there were on average 33% (10%) surviving beyond the end of follow-up.

The results are displayed in Tables 1 and 2 below. The random number generator used the same initial seed for the runs marked a and different seeds for the runs marked b and c. Runs with the same initial seed have identical failure times.

Table 1. Observed Coverage Probabilities with 500 Simulations using Exponential Censoring ($\beta = 0.3$) – 28% Censoring.

d	$n = 1000^a$		$n = 4000^b$	
	Covariate 1	Covariate 2	Covariate 1	Covariate 2
8	.976	.982	.982	.966
16	.972	.980	.978	.964
64	.958	.970	.966	.960

Table 2. Observed Coverage Probabilities with 500 Simulations using Exponential Censoring ($\beta = 1.5$) – 68% Censoring.

d	$n = 1000^a$		$n = 4000^c$	
	Covariate 1	Covariate 2	Covariate 1	Covariate 2
8	.978	.982	.982	.982
16	.972	.980	.980	.984
64	.958	.974	.974	.980

Observe from Tables 1 and 2 that the bands appear to be conservative but as d increases the coverage probabilities, on the whole, get closer to their nominal value of .95. This is true for both covariates under both light (28%) and heavy (68%) censoring. On the other hand, in all the simulations we carried out, the coverage probabilities for the pointwise confidence intervals were very close (not significantly different) to .95, a typical case being given in Table 3 below (see column 5).

Table 3. True and Average Estimated Cumulative Hazard Function, Standard Deviation of the Simulated Estimates and confidence Interval Coverage Probabilities with 500 Simulations, 28% Censoring ($\beta = 0.3$), $n = 1000$, $d = 8$, Covariate 1.

Time	True integrated hazard	Average of simulated estimates	Standard deviation of simulated estimates	Proportion of confidence intervals containing the true value
.125	.125	.1236	.026	.940
.25	.25	.2513	.038	.940
.375	.375	.3781	.049	.954
.5	.5	.5024	.060	.948
.625	.625	.6243	.071	.952
.75	.75	.7501	.082	.960
.875	.875	.8736	.096	.948
1.0	1.0	1.0010	.113	.948

Also observe from Table 3 that the estimators appear to be unbiased, in support of the heuristic arguments in section 2.

Figures 1-4 contain plots of the estimators under light (28%) and heavy (68%) censoring for $d=8$ and $d=64$. The sample size n was set to 2000 and the random number generator used the same initial seed for all runs, so the failure times used in each run are identical.

[Insert Figures 1-4 here]

As expected, the bands are wider under the heavier censoring; compare Figures 3 and 4 with Figures 1 and 2. Also, it appears that the estimator which uses $d=8$ (in Figures 1 and 3) is a smoothed version of the estimator which uses $d=64$ (in Figures 2 and 4). Although the estimators are adequate in each case, notice that under heavy censoring the estimator which uses $d=64$ (in Figure 4) oscillates considerably when t is close to 1. This is due to the very small number of observed failures (averaging less than 5) for each of

the sieve intervals in this region. A more flexible procedure would be to merge adjoining intervals which contain very few observed failures. For contrast, in Figure 2 notice that the oscillation is negligible. This is because of the lower censoring rate which gives an average of more than 10 observed failures per interval near $t=1$.

4. APPLICATION TO THE ANALYSIS OF CANCER MORTALITY AMONG JAPANESE ATOMIC BOMB SURVIVORS

The Radiation Effects Research Foundation (RERF) in Hiroshima, Japan, has followed since 1950 a group of over 100,000 atomic bomb survivors. Data on these survivors is the primary source of information on the epidemiologic effects of ionizing radiation. The National Research Council (1980) report and the report by Kato and Schull (1982) contain detailed analyses of these data. However, as noted by Pierce and Preston (1984), a difficulty with the methods used in these reports is that they do not make explicit allowance for temporal variation in risks; they simply average the risk over the follow-up period.

Pierce and Preston (1985) analyzed cancer mortality among the atomic bomb survivors using a parametric additive risk model which they called the excess risk model. They found that background and excess rates of cancer mortality vary markedly with age at exposure and time since exposure.

The approach taken here is to fit a nonparametric additive risk model of the form (1.1) for each of 4 cohorts defined by the age at exposure intervals 0-9, 10-19, 20-34 and 35-49 years of age at time of bomb. The time variable t is time since exposure, ranging from 5 to 37 years over the follow-up period from 1950 to 1982. The only censoring prior to the end of follow-up is that due to other (non-cancer) causes of death. Summary information for each cohort is give in Table 4.

Table 4. Cohort Sizes, Summary Mortality Figures and Censoring Prior to End of Follow-up.

Age at exposure	Cohort size*	Deaths due to Cancer excluding Leukemia	Deaths from all causes	Censoring
0-9	18416	93	728	87%
10-19	19242	349	1715	80%
20-34	17694	949	3075	69%
35-49	20916	2788	11234	75%

* Approximate, having been estimated from the grouped data.

Three covariates are used. For individual i they are: Y_{i1} = indicator (male), Y_{i2} = indicator (female), Y_{i3} = dose (in units of 100 rads). The hazard functions α_1 and α_2 are the background cancer mortality rates for males and females respectively. The third hazard function α_3 is the excess cancer mortality rate per 100 rads of radiation exposure.

In the data provided to us by the RERF time since exposure is grouped into eight 4-year intervals: 5-9, ..., 33-37 years since exposure. It is natural to use these eight time intervals as the partition of the follow-up period in the evaluation of the sieve estimators. As in Pierce and Preston (1985) the dose variable is taken as the midpoint of one of the six dose groups: 0, 1-50, 50-100, 100-200, 200-300, > 300 rads with dose=400 for dose > 300. Average doses for the dose groups are not used because of current reevaluation of the dosimetry. Also, the analysis is limited to the epithelial cancer mortality, in which leukemia mortality is excluded.

Figures 5-8 indicate our estimates and 95% confidence intervals and bands for the background and excess cumulative cancer mortality rates as functions of time since exposure. The estimates of Pierce and Preston are also plotted. All estimates are given in units of deaths per 1000 persons at risk. Pierce and Preston's parametric model for the cancer mortality rate $\lambda(t)$ is given by

$$\lambda(t) = \exp \{ \nu_{0e} + \nu_{1e} + \nu_{2e} \log(e + t) \} + \beta d \exp \{ \rho_{0e} + \rho_{1e} \log(t) \}, \quad (4.1)$$

where e = midpoint of age at exposure interval, s = sex, and d = dose. The first and second terms in (4.1) represent the background and excess cancer mortality rates which Pierce and Preston fitted by maximum likelihood. We have integrated these to provide a comparison with the integrated histogram sieve estimates.

[Insert Figures 5-8 here]

From Figures 5-8 we see that there is a significant dose effect (i.e. the band for dose does not contain the zero function) for all cohorts. This is despite the fact that the confidence bands with eight sieve intervals are conservative according to our simulation results in section 3. In interpreting Figures 5-8 bear in mind that the vertical scales are different for each cohort. Despite appearances, the dose effect bands have roughly the same width for each cohort.

On the whole our estimates are in agreement with those of Pierce and Preston (1985). We also observe that the relative risk (dose effect vs. background) decreases sharply with age at exposure. Our estimates for the dose effect are somewhat higher than Pierce and Preston's, but their estimates are still within our 95% bands. However, there is a significant difference between our estimates and Pierce and Preston's for the female background mortality rate in all but the 0-9 years of age cohort.

The data used for the analyses in this section are contained in the file R1OCANCR.DAT which is described in the Life Span Study Report 10, Part I published by RERF. The file was supplied to us by RERF which is a private foundation that is funded equally by the Japanese Ministry of Health and Welfare and the U.S. Department of Energy through the U.S. National Academy of Sciences. The conclusions reached in this paper are those of the authors and do not necessarily reflect the scientific judgement of RERF or its funding agencies.

5. ASYMPTOTIC RESULTS

In this section we give a general formulation of the asymptotic results which underlie the confidence bands in section 2. Definitions of the martingale concepts used in this section can be found in the review paper of Andersen and Borgan (1985).

First we look at the basic counting processes and martingales involved in the random censorship model. Let $N_i(t)$ be the indicator of an uncensored failure for subject i prior to time t ,

$$N_i(t) = I(\tilde{X}_i \leq t, \delta_i = 1).$$

Aalen (1978) showed that the counting process $N_i(t)$ can be written in the form

$$N_i(t) = \sum_{j=1}^p \int_0^t \alpha_j(s) Y_{ij}(s) ds + M_i(t),$$

where $Y_{ij}(t) = Y_{ij}I(\tilde{X}_i \geq t)$ and M_i is a square integrable martingale with respect to the filtration $\mathcal{F}_t = \sigma(N_i(s), Y_{ij}(s+), 0 \leq s \leq t, i \geq 1, j = 1, \dots, p)$. Note that no two of the counting processes N_1, \dots, N_n jump simultaneously.

More generally, let $N(t) = (N_1(t), \dots, N_n(t))'$, $t \in [0, 1]$ be a multivariate counting process with respect to a right-continuous filtration (\mathcal{F}_t) , i.e. N is adapted to the filtration and has components N_i which have sample paths which are non-decreasing, right-continuous step functions, zero at time zero, and with jumps of unit-size. Moreover, suppose that no two components jump simultaneously. Let Λ be the compensator of N , so that $N = \Lambda + M$ where $M = (M_1, \dots, M_n)'$ and M_1, \dots, M_n are local martingales. Aalen (1980) introduced the model

$$\Lambda(t) = \int_0^t Y(s) \alpha(s) ds \quad (5.1)$$

where $\alpha = (\alpha_1, \dots, \alpha_p)'$ is a vector of unknown nonrandom integrable functions, $Y(t) = (Y_{ij}(t))$ is an $n \times p$ matrix of covariate processes, with $Y_{ij}(t)$ representing the j th covariate for the i th subject at time t . The covariate processes are assumed to be predictable and locally bounded (which is the case for instance if they are left-continuous with right-hand limits and adapted). This model allows much more general censoring mechanisms (quite arbitrary apart from being predictable) than the one considered in section 2. Unlike Aalen we do not assume, unless explicitly stated, that $EN_i(1) < \infty$ (in which case M_i is a square integrable martingale), only that $N_i(1) < \infty$ almost surely.

Aalen (1980) discussed estimators A^* of A of the form

$$A^*(t) = \int_0^t Y^-(s) dN(s), \quad (5.2)$$

where $Y^-(s)$ is a predictable generalized inverse of $Y(s)$. In the case $p = 1$, the choice

$$(Y^-(s))_{1i} = \left(\sum_{k=1}^n Y_{k1}(s) \right)^{-1}, \quad i = 1, \dots, n$$

(where $1/0 \equiv 0$) gives the Nelson-Aalen estimator for which a general asymptotic theory was derived by Aalen (1978). The powerful martingale techniques used in this theory allow the conventional i.i.d. structure to be superceded; only very weak asymptotic stability and Lindeberg conditions need to be imposed on the (single) covariate, see Andersen and Borgan (1985, p. 136).

The case $p > 1$ is more complicated. First, it is not clear which generalized inverse Y^- of Y to use. Aalen (1980, Remark 1) considered

$$Y^-(s) = (Y'(s)Y(s))^{-1}Y'(s) \quad (5.3)$$

which he motivated by a formal least squares principle and noted that the resulting estimator

$$\tilde{A}(t) = \int_0^t (Y'(s)Y(s))^{-1}Y'(s)dN(s) \quad (5.4)$$

probably gives reasonable (though not optimal) estimates of A . Second, the problem of finding the asymptotic distribution of the *general* estimator of the form (5.2) is intractable.

5.1. Aalen's least squares estimator

Despite the difficulties alluded to above it is possible to obtain the asymptotic distribution of \tilde{A} in a straightforward way by applying Rebolledo's Central Limit Theorem for local square integrable martingales in the form given by Andersen and Gill (1982, Theorem I.2). Let $D[0,1]^p$ be the product of p copies of $D[0,1]$ and equip it with the Skorohod product topology. Denote

$$\tilde{K}_{jk}(t) = \frac{1}{n} \sum_{i=1}^n Y_{ij}(t)Y_{ik}(t) \quad (5.5)$$

$$\tilde{R}_{jk\ell}(t) = \frac{1}{n} \sum_{i=1}^n Y_{ij}(t)Y_{ik}(t)Y_{i\ell}(t) \quad (5.6)$$

CONDITIONS

(A1) (Asymptotic stability). For $j, k, \ell = 1, \dots, p$ there exist bounded functions K_{jk} and $R_{jk\ell}$ defined on $[0, 1]$ such that

$$\sup_{t \in [0, 1]} |\tilde{K}_{jk}(t) - K_{jk}(t)| \xrightarrow{P} 0$$

$$\sup_{t \in [0, 1]} |\tilde{R}_{jk\ell}(t) - R_{jk\ell}(t)| \xrightarrow{P} 0.$$

(A2) (Lindeberg condition). For each $j = 1, \dots, p$

$$n^{-\frac{1}{2}} \sup_{i, t} |Y_{ij}(t)| \xrightarrow{P} 0.$$

(A3) (Asymptotic regularity condition). The $p \times p$ matrix function $K(t) = (K_{jk}(t))$ in (A1) satisfies

$$\inf_{t \in [0, 1]} \det K(t) > 0.$$

THEOREM 5.1. *Under conditions (A1)-(A3)*

$$\sqrt{n}(\tilde{A} - A) \xrightarrow{D} m \text{ in } D[0, 1]^p$$

where m is a p -variate continuous Gaussian martingale with $m(0) = 0$ and covariance functions

$$\text{Cov}(m_j(t), m_k(t)) = G_{jk}(t)$$

where

$$G_{jk}(t) = \sum_{u=1}^p \sum_{v=1}^p \sum_{w=1}^p \int_0^t R_{uvw}(s) (K^{-1}(s))_{jv} (K^{-1}(s))_{kw} \alpha_u(s) ds. \quad (5.7)$$

The following theorem shows that

$$\tilde{G}_{jk}(t) = \sum_{u=1}^p \sum_{v=1}^p \sum_{w=1}^p \int_0^t \tilde{R}_{uvw}(s) (\tilde{K}^{-1}(s))_{jv} (\tilde{K}^{-1}(s))_{kw} d\tilde{A}_u(s) \quad (5.8)$$

is a uniformly consistent estimator of the covariance matrix of the limiting Gaussian martingale. In the survival analysis context \tilde{G}_{jj} is given by (2.6).

THEOREM 5.2. *Under conditions (A1)-(A3), for each $j, k = 1, \dots, p$*

$$\sup_{t \in [0,1]} |\tilde{G}_{jk}(t) - G_{jk}(t)| \xrightarrow{P} 0.$$

Combining Theorems 5.1 and 5.2 with standard results from Billingsley (1967, Theorems 4.1 and 5.1) shows that (2.5) is an asymptotic $100(1 - \alpha)\%$ confidence band for A_j . The transformation to the Brownian bridge which is used in deriving this confidence band is described by Andersen and Borgan (1985, p. 114).

The Lindeberg condition (A2) holds if the covariates are bounded by random variables having a bounded r th moment for some $r > 2$, c.f. the discussion of Andersen and Gill (1982, p.1110) in the context of Cox's proportional hazards model.

The following theorem gives conditions under which (A1)-(A3) hold in the i.i.d. case in which the $(N_i, Y_{i1}, \dots, Y_{ip})$ are i.i.d. replicates of $(N_1, Y_{11}, \dots, Y_{1p})$.

THEOREM 5.3. *In the i.i.d. case with Y left-continuous with right-hand limits, conditions (A1)-(A3) are satisfied if for each $j = 1, \dots, p$*

$$E \left(\sup_{t \in [0,1]} |Y_{1j}^3(t)| \right) < \infty \quad (5.9)$$

and the $p \times p$ matrix $K(t)$, now defined by $K_{jk}(t) = EY_{1j}(t)Y_{1k}(t)$, satisfies condition (A3).

5.2. Integrated histogram sieve estimator

The histogram sieve estimator is defined in the setting of the counting process model (5.1) by $\hat{\alpha}(t) = \alpha_r$, for $t \in I_r$, where $\hat{\alpha}_r = (\hat{\alpha}_{rj})$ is the $p \times 1$ vector give by $\hat{\alpha}_r = D_r^{-1}C_r$, and C_r is the $p \times 1$ vector

$$C_r = \int_{I_r} Y'(s) dN(s),$$

D_r is the $p \times p$ matrix

$$D_r = \int_{I_r} Y'(s)Y(s)ds.$$

Note the similarity between the integrated histogram sieve estimator

$$\hat{A}(t) = \int_0^t \hat{\alpha}(s)ds$$

and Aalen's least squares estimator. However \hat{A} is not (\mathcal{F}_t) - adapted and the derivation of its asymptotic distribution theory is much more difficult. The following result of McKeague (1987) gives the asymptotic distribution of \hat{A} .

CONDITIONS.

(C1) (Asymptotic stability). For $j, k, \ell = 1, \dots, p$ there exist function K_{jk} and $R_{jk\ell}$ defined on $[0, 1]$ such that

$$\sup_{t \in [0, 1]} |\tilde{K}_{jk}(t) - K_{jk}(t)| = O_P(n^{-\frac{1}{2}})$$

$$\sup_{t \in [0, 1]} |\tilde{R}_{jk\ell}(t) - R_{jk\ell}(t)| \xrightarrow{L^1} 0.$$

(C2) (Lindeberg condition). Same as (A2).

(C3) (Asymptotic regularity condition). The $p \times p$ matrix $K(t) = (K_{jk}(t))$ is nonsingular for all $t \in [0, 1]$.

(C4) The functions $\alpha_j, K_{jk}, R_{jk\ell}$ $j, k, \ell = 1, \dots, p$ are Lipschitz.

THEOREM 5.4. *If conditions (C1)–(C4) hold, $EN_i(1) < \infty$ for all $i \geq 1$, $\sqrt{n} \max(\ell_1, \dots, \ell_d) \rightarrow 0$, and $n \min(\ell_1, \dots, \ell_d) \rightarrow \infty$ then*

$$\sqrt{n}(\hat{A} - A) \xrightarrow{D} m \text{ in } C[0, 1]^p$$

where m is the p -variate continuous Gaussian martingale of Theorem 5.1.

The appropriate (sieve based) estimator of the covariance matrix of the limiting Gaussian martingale is given by

$$\hat{G}_{jk}(t) = \sum_{u=1}^p \sum_{v=1}^p \sum_{w=1}^p \int_0^t \hat{\alpha}_u(s) \hat{R}_{uvw}(s) (\hat{K}^{-1}(s))_{jv} (\hat{K}^{-1}(s))_{kw} ds$$

where $\hat{K}(t)$ is the $p \times p$ matrix with entries

$$\hat{K}_{jk}(t) = \frac{1}{n\ell_r} \sum_{i=1}^n \int_{I_r} Y_{ij}(s) Y_{ik}(s) ds, \text{ for } t \in I_r$$

and

$$\hat{R}_{uvw}(t) = \frac{1}{n\ell_r} \sum_{i=1}^n \int_{I_r} Y_{iu}(s)Y_{iv}(s)Y_{iw}(s)ds, \text{ for } t \in I_r.$$

In the survival analysis context \hat{G}_{jj} simplifies to (2.2). It can be shown, under the hypotheses of Theorem 5.4, that \hat{G}_{jj} is a uniformly consistent estimator of G_{jj} (McKeague, 1987). It follows that (2.1) is an asymptotic $100(1 - \alpha)\%$ confidence band for A_j .

5.3. Weighted least squares estimators

Neither Aalen's least squares estimator nor the integrated histogram sieve estimator is an optimal estimator. In the case of a single covariate the Nelson-Aalen estimator is optimal. However, if *weighted* least squares is used instead of ordinary least squares, it is possible to improve the performance of Aalen's least squares and the integrated histogram sieve estimators. Specifically, consider the weighted least squares estimator

$$\tilde{\tilde{A}}(t) = \int_0^t (Y'(s)W(s)Y(s))^{-1}Y'(s)W(s)dN(s) \quad (5.10)$$

where $W(t)$ is the $n \times n$ diagonal matrix having i th diagonal entry

$$W_i(t) = \left(\sum_{j=1}^p \tilde{\alpha}_j(i)Y_{ij}(t) \right)^{-1}, \quad i = 1, \dots, n$$

where $\tilde{\alpha}_j$ is a predictable estimator of α_j and $1/0 \equiv 0$. Note that $\tilde{\tilde{A}}$ is an estimator of the form (5.2). In the case of a single covariate $\tilde{\tilde{A}}$ coincides with the Nelson-Aalen estimator and no companion estimator $\tilde{\alpha}_1$ is needed. For more than one covariate a predictable estimator $\tilde{\alpha}_j$ for α_j can be constructed by kernel smoothing of Aalen's (ordinary) least squares estimator \tilde{A} as follows. Let

$$\tilde{\alpha}_j(t) = \frac{1}{b_n} \int_0^1 K\left(\frac{t-s}{b_n}\right)d\tilde{A}_j(s),$$

where K is a continuous function having support $[0,1]$ and integral 1, and $b_n > 0$ is a bandwidth parameter which tends to zero as $n \rightarrow \infty$. Under conditions on α_j, k and b_n uniform consistency of $\tilde{\alpha}_j$ can be shown using similar techniques to those of Ramlau-Hansen (1983) for the Nelson-Aalen estimator. Finally, along the lines of the proof of Theorem

5.1, it can be shown under mild regularity conditions that $\sqrt{n}(\tilde{A} - A)$ converges weakly in $D[0, 1]^p$ to a p -variate Gaussian martingale m with $m(0) = 0$ and covariance matrix

$$\text{Cov}(m(t), m(t)) = \int_0^t V^{-1}(s) ds$$

where (in the i.i.d. case) $V(s)$ is the $p \times p$ matrix with entries

$$V_{jk}(t) = E \left[\frac{Y_{1j}(t) Y_{1k}(t)}{\sum_{\ell=1}^p \alpha_{\ell}(t) Y_{1\ell}(t)} \right].$$

The weighted version of the histogram sieve estimator and its integrated counterpart can be developed in an analogous fashion. The weighted least squares approach will be treated at length in a subsequent paper.

APPENDIX

PROOF OF THEOREM 5.1. First split $\sqrt{n}(\tilde{A}(t) - A(t))$ into four parts using (5.1), (5.4) and (5.5):

$$\sqrt{n}(\tilde{A}(t) - A(t)) = \frac{1}{\sqrt{n}} \int_0^t K^{-1}(s) Y'(s) dM(s) \quad (A.1)$$

$$+ \frac{1}{\sqrt{n}} \int_0^t (J(s) - 1) K^{-1}(s) Y'(s) dM(s) \quad (A.2)$$

$$+ \frac{1}{\sqrt{n}} \int_0^t J(s) [\tilde{K}^{-1}(s) - K^{-1}(s)] Y'(s) dM(s) \quad (A.3)$$

$$+ \sqrt{n} \int_0^t (J(s) - 1) dA(s) \quad (A.4)$$

where

$$J(t) = I(\tilde{K}(t) \text{ is invertible}). \quad (A.5)$$

The local martingales M_1, \dots, M_n are local square integrable martingales on the time interval $[0, 1]$ and their predictable quadratic variation processes are given by

$$\langle M_i, M_i \rangle(t) = \int_0^t \lambda_i(s) ds \quad \text{and} \quad \langle M_i, M_j \rangle(t) = 0, \quad i \neq j \quad (A.6)$$

where

$$\lambda_i(t) = \sum_{j=1}^p \alpha_j(t) Y_{ij}(t).$$

Now consider (A.1) which can be written as $X^{(n)}(t) = (X_1^{(n)}(t), \dots, X_p^{(n)}(t))'$ where

$$X_j^{(n)}(t) = \sum_{i=1}^n \int_0^t H_{ij}^{(n)}(s) dM_i(s)$$

$$H_{ij}^{(n)}(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^p (K^{-1}(t))_{jk} Y_{ik}(t).$$

Condition (A3) implies that the functions $(K^{-1}(t))_{jk}$ are bounded so that $X_j^{(n)}$ is a local square integrable martingale. Also, from (A.6)

$$\langle X_j^{(n)}, X_k^{(n)} \rangle(t) = \int_0^t \sum_{i=1}^n H_{ij}^{(n)}(s) H_{ik}^{(n)}(s) \lambda_i(s) ds$$

so that using the asymptotic stability condition (A1), for all j, k and t

$$\langle X_j^{(n)}, X_k^{(n)} \rangle(t) \xrightarrow{P} G_{jk}(t)$$

as $n \rightarrow \infty$. Next, the boundedness of the functions $(K^{-1}(t))_{jk}$ and condition (A2) imply the following Lindeberg condition: for all j and $\varepsilon > 0$

$$\int_0^1 \sum_{i=1}^n H_{ij}^{(n)}(t)^2 \lambda_i(t) I(|H_{ij}^{(n)}(t)| > \varepsilon) dt \xrightarrow{P} 0$$

as $n \rightarrow \infty$. By Rebolledo's Central Limit Theorem for local square integrable martingales in the form given by Andersen and Gill (1982, Theorem I.2) it follows that $X^{(n)} \xrightarrow{D} m$ in $D[0, 1]^p$.

Note that by conditions (A1) and (A3),

$$P(\tilde{K}(t) \text{ is invertible for all } t \in [0, 1]) \rightarrow 1 \quad (A.7)$$

as $n \rightarrow \infty$. This shows that terms (A.2) and (A.4) converge uniformly to zero in probability as $n \rightarrow \infty$.

Finally consider (A.3) whose j th component is given by

$$Z_j^{(n)}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t J(s) \sum_{k=1}^p [\tilde{K}^{-1}(s) - K^{-1}(s)]_{jk} Y_{ik}(s) dM_i(s)$$

which is a local square integrable martingale. By Lengart's (1977) inequality, for each $\varepsilon > 0, \eta > 0$

$$P\left(\sup_{t \in [0,1]} |Z_j^{(n)}(t)| > \varepsilon\right) \leq \frac{\eta}{\varepsilon^2} + P(\langle Z_j^{(n)}, Z_j^{(n)} \rangle > (1) \geq \eta). \quad (\text{A.8})$$

Now evaluate $\langle Z_j^{(n)}, Z_j^{(n)} \rangle$ using (A.6) to obtain

$$\begin{aligned} \langle Z_j^{(n)}, Z_j^{(n)} \rangle > (1) &= \frac{1}{n} \sum_{i=1}^n \int_0^1 J(s) \left\{ \sum_{k=1}^p [\tilde{K}^{-1}(t) - K^{-1}(t)]_{jk} Y_{ik}(t) \right\}^2 \lambda_i(t) dt \\ &\leq p \sum_{k=1}^p \left\{ \sup_{t \in [0,1]} |[\tilde{K}^{-1}(t) - K^{-1}(t)]_{jk}| \right\} \left\{ \frac{1}{n} \sum_{i=1}^n \int_0^1 Y_{ik}^2(s) \lambda_i(s) ds \right\}. \quad (\text{A.9}) \end{aligned}$$

By the componentwise continuity of the matrix inverse operation and conditions (A1), (A3) we have for the first term in (A.9)

$$\sup_{t \in [0,1]} |[\tilde{K}^{-1}(t) - K^{-1}(t)]_{jk}| \xrightarrow{P} 0. \quad (\text{A.10})$$

Also, using condition (A1), for the second term in (A.9)

$$\frac{1}{n} \sum_{i=1}^n \int_0^1 Y_{ik}^2(s) \lambda_i(s) ds \xrightarrow{P} \sum_{j=1}^p \int_0^1 R_{kkj}(t) \alpha_j(t) dt. \quad (\text{A.11})$$

Combining (A.8)-(A.11) we conclude that (A.3) converges uniformly to zero in probability as $n \rightarrow \infty$. This completes the proof of the theorem. ■

PROOF OF THEOREM 5.2. Fix j and k . Let $L(t)$ and $\tilde{L}(t)$ be the $1 \times p$ vectors with components

$$\begin{aligned} L_u(t) &= \sum_{u=1}^p \sum_{w=1}^p R_{uw}(t) (K^{-1}(t))_{jv} (K^{-1}(t))_{kw} \\ \tilde{L}_u(t) &= \sum_{v=1}^p \sum_{w=1}^p \tilde{R}_{uw}(t) (\tilde{K}^{-1}(t))_{jv} (\tilde{K}^{-1}(t))_{kw}. \end{aligned}$$

Using (5.1), (5.4) and (5.5) we can write

$$\tilde{G}_{jk}(t) - G_{jk}(t) = \frac{1}{n} \int_0^1 \tilde{L}(s) \tilde{K}^{-1}(s) Y'(s) dM(s) \quad (\text{A.12})$$

$$+ \int_0^t J(s) (\tilde{L}(s) - L(s)) dA(s) \quad (\text{A.13})$$

$$+ \int_0^t (J(s) - 1) L(s) dA(s). \quad (\text{A.14})$$

The term (A.14) is dealt with using (A.7). From (A.10) and the asymptotic stability condition (A1)

$$\sup_{t \in [0,1]} |\tilde{L}_u(t) - L(t)| \xrightarrow{P} 0,$$

which deals with (A.13). Finally, using Lengart's inequality it is seen that (A.12) also converges uniformly in probability to zero. ■

PROOF OF THEOREM 5.3. Apply the Strong Law of Large Numbers in $D[0,1]$ (Rao, 1963) in the reversed time direction. ■

Acknowledgment. The authors thank Dr. David G. Hoel for his help in obtaining the data used in Section 4.

References

- Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* 6, 701-726.
- Aalen, O.O. (1980). A model for nonparametric regression analysis of counting processes. *Lecture Notes in Statistics* 2, 1-25. Springer-Verlag, New York.
- Andersen, P.K., Borgan, O. (1985). Counting process model for life history data: a review (with discussion). *Scand. J. Statist.* 12, 97-158.
- Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* 10, 1100-1120.
- Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- Breslow, N.E. (1986). Cohort analysis in epidemiology. In A.C. Atkinson and S.E. Fienberg (eds.), *A Celebration of Statistics: The ISI Centenary Volume*. Springer-Verlag, New York, 109-143.
- Buckley, J.D. (1984). Additive and multiplicative models for relative survival rates. *Biometrics* 40, 51-62.
- Cox, D.R. (1972). Regression models with life tables (with discussion). *J. Roy. Statist. Soc. B* 34, 187-220.
- Grenander, U. (1981). *Abstract Inference*. Wiley, New York.
- Hall, W.J. and Wellner, J.A. (1980). Confidence bands for a survival curve with censored data. *Biometrika* 67, 133-143.
- Kato, H. and Schull, W.J. (1982). Studies of the mortality of A-bomb survivors 7. Mortality, 1950-1978: Part I. Cancer mortality. *Radiation Research* 90, 395-432.
- Lenglart, E. (1977). Relation de domination entre deux processus. *Ann. Inst. H. Poincaré* 13, 171-179.
- McKeague, I.W. (1987). Nonparametric inference in additive risk models for counting processes. To appear in *Stochastic Process. Appl.*
- National Research Council (1980). *The Effects on Populations of Exposure to Low Levels of Ionizing Radiation*. National Academy of Sciences, Washington, D.C.

O'Neill, T.J. (1986). Inconsistency of the misspecified proportional hazards model. *Statistics and Probability Letters* 4, 219-222.

Pierce, D.A. and Preston, D.L. (1984). Hazard function modelling for dose-response analysis of cancer incidence in the A-bomb survivor cohort. *Atomic Bomb Survivor Analysis: Utilization and Analysis*. Proceedings of a Conference of the SIAM Institute for Mathematics and Society, pp. 51-66. SIAM, Philadelphia.

Pierce, D.A. and Preston, D.L. (1985). Analysis of cancer mortality in the atomic bomb survivor cohort. Technical Report 108, Department of Statistics, Oregon State University.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* 11, 453-466.

Rao, R.R. (1963). The law of large numbers for $D[0, 1]$ -valued random variables. *Theor. Probab. Appl.* 8, 70-74.

Key to Figures 1-8

Thick lines = integrated histogram sieve estimates.

Regular dashed lines = 95% pointwise confidence limits.

Thin lines = 95% confidence bands.

Irregular dashed lines = true integrated hazard functions (in Figures 1-4).

Lines marked with boxes = estimates of Pierce and Preston (in Figures 5-8)

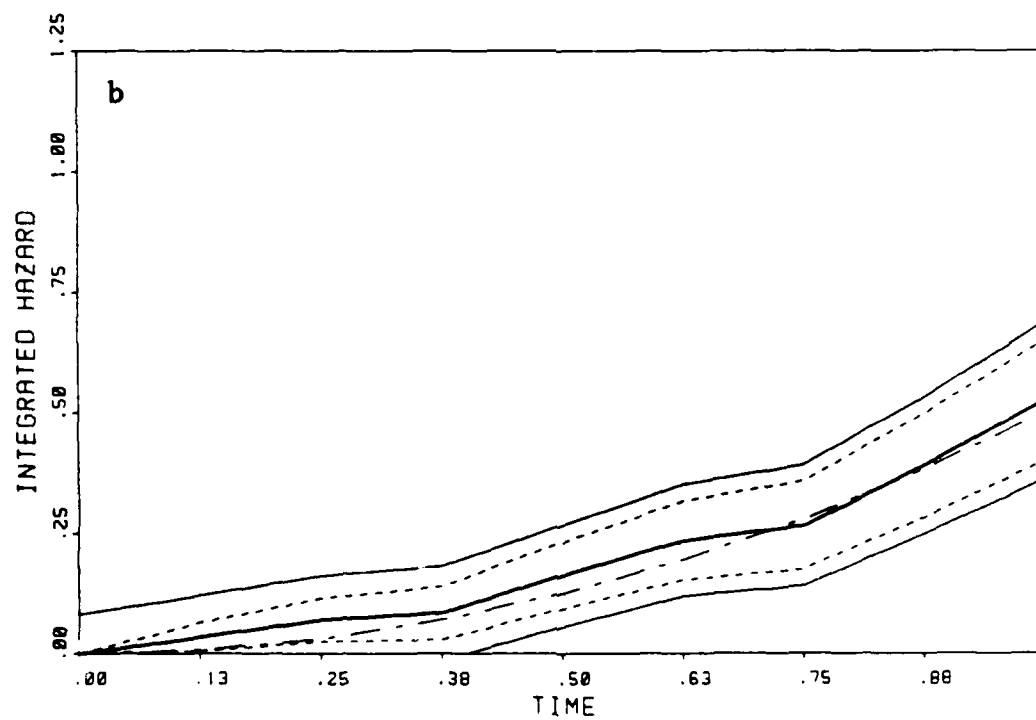
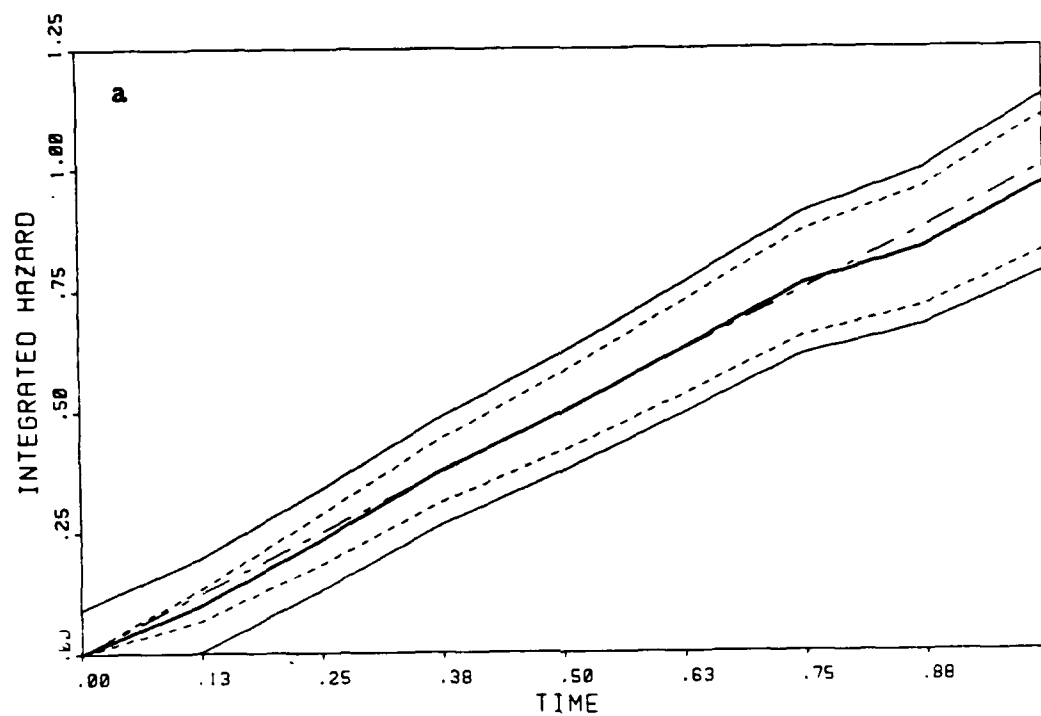


Figure 1. Light censoring (28%), $n = 2000$, $d = 8$. (a) Covariate 1. (b) Covariate 2.

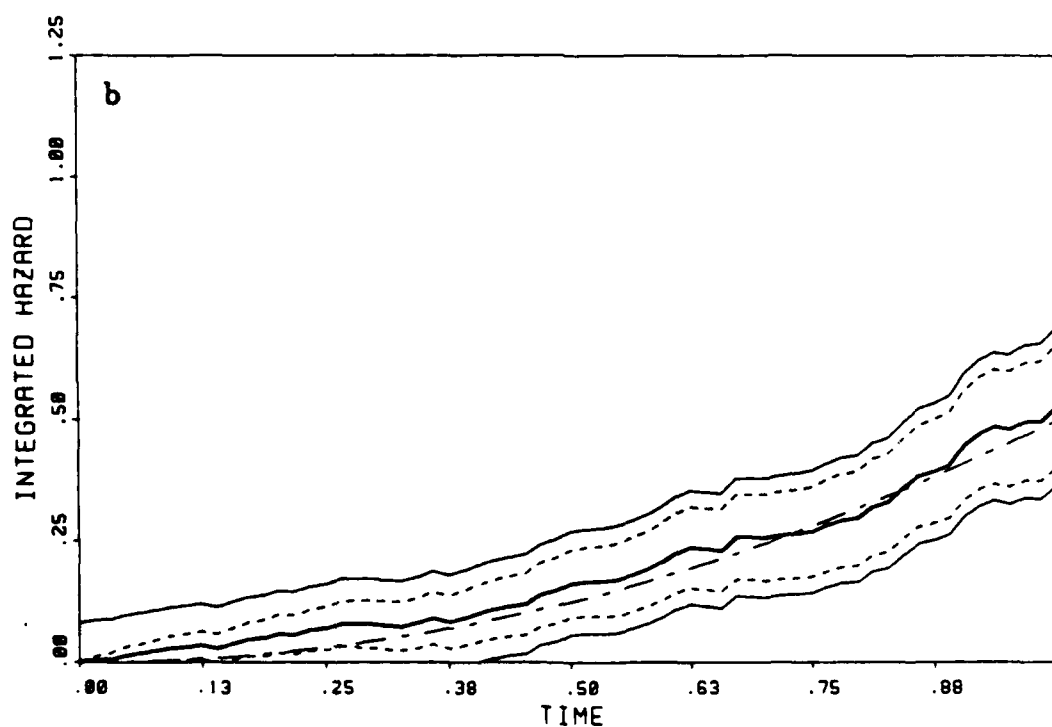
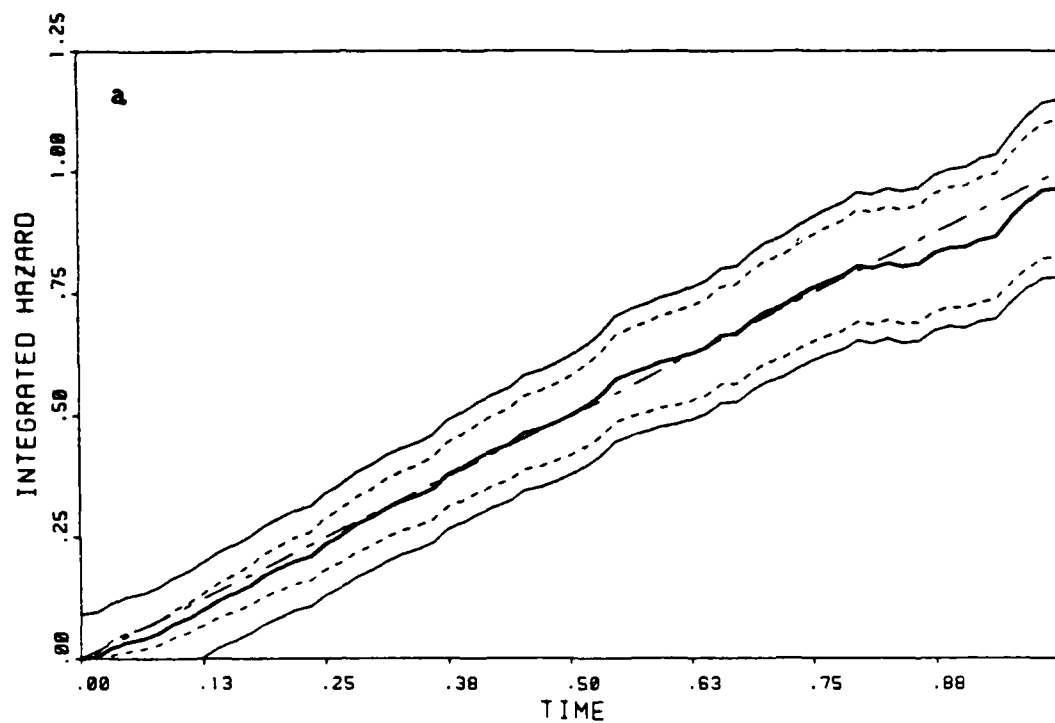


Figure 2. Light censoring (28%), $n = 2000$, $d = 64$. (a) Covariate 1. (b) Covariate 2.

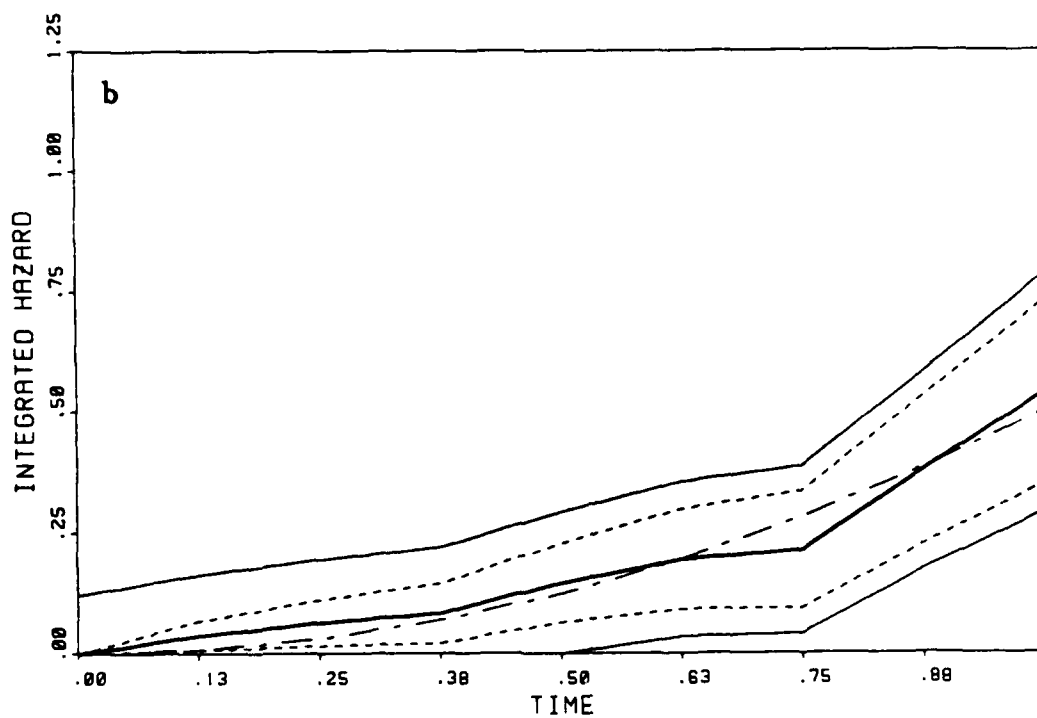
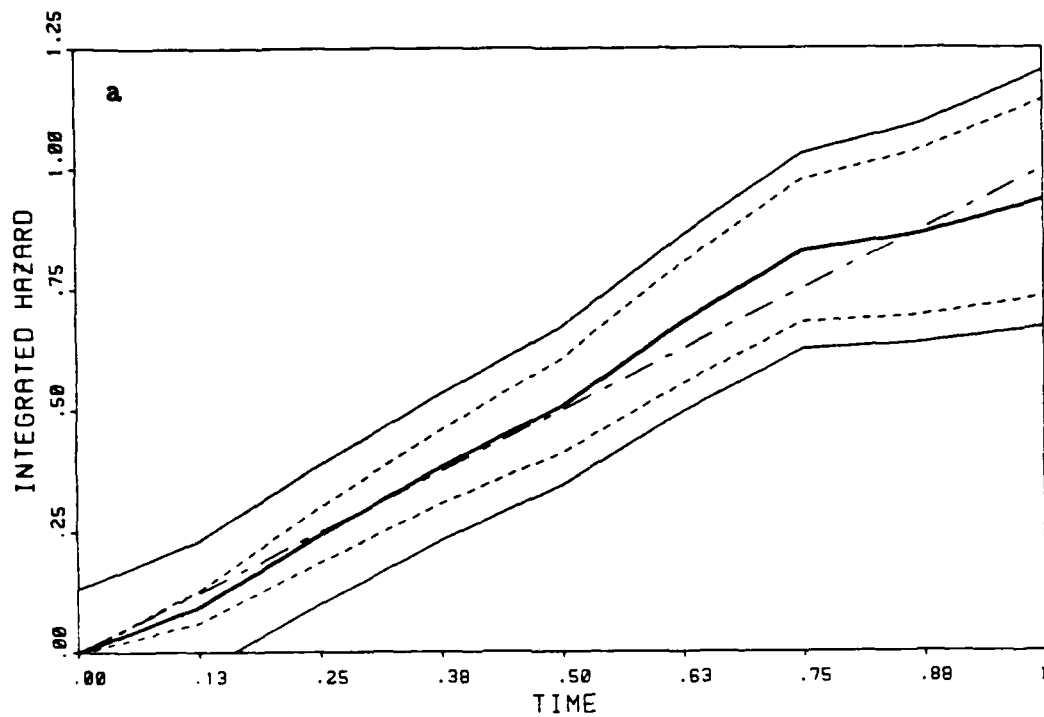


Figure 3. Heavy censoring (68%), $n = 2000$, $d = 8$. (a) Covariate 1. (b) Covariate 2.

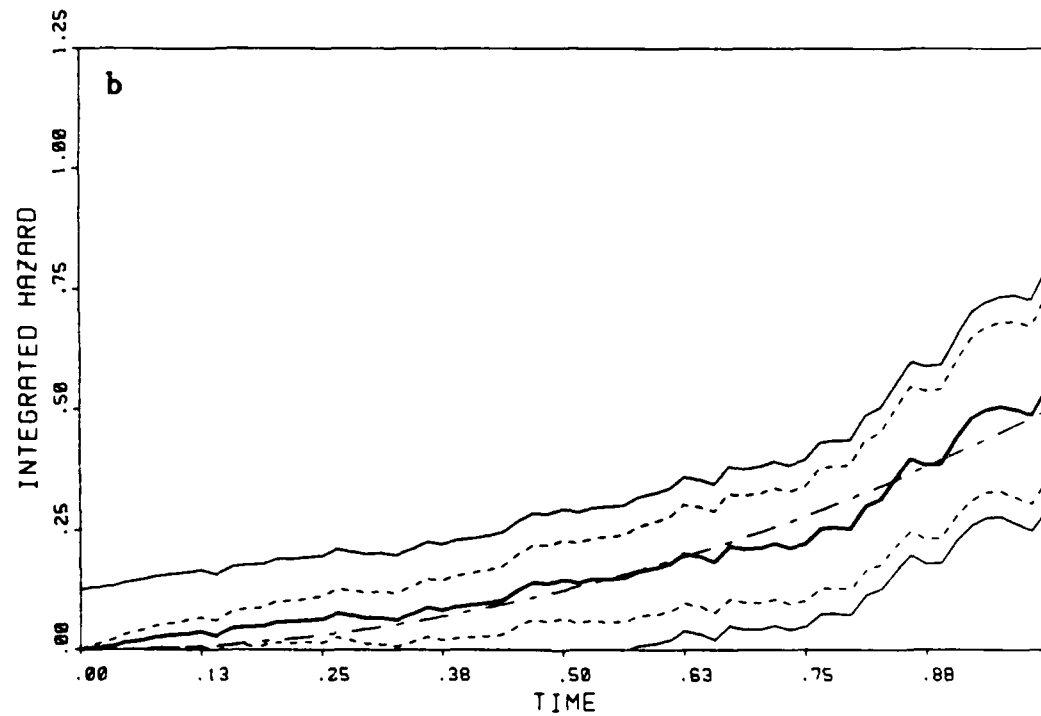
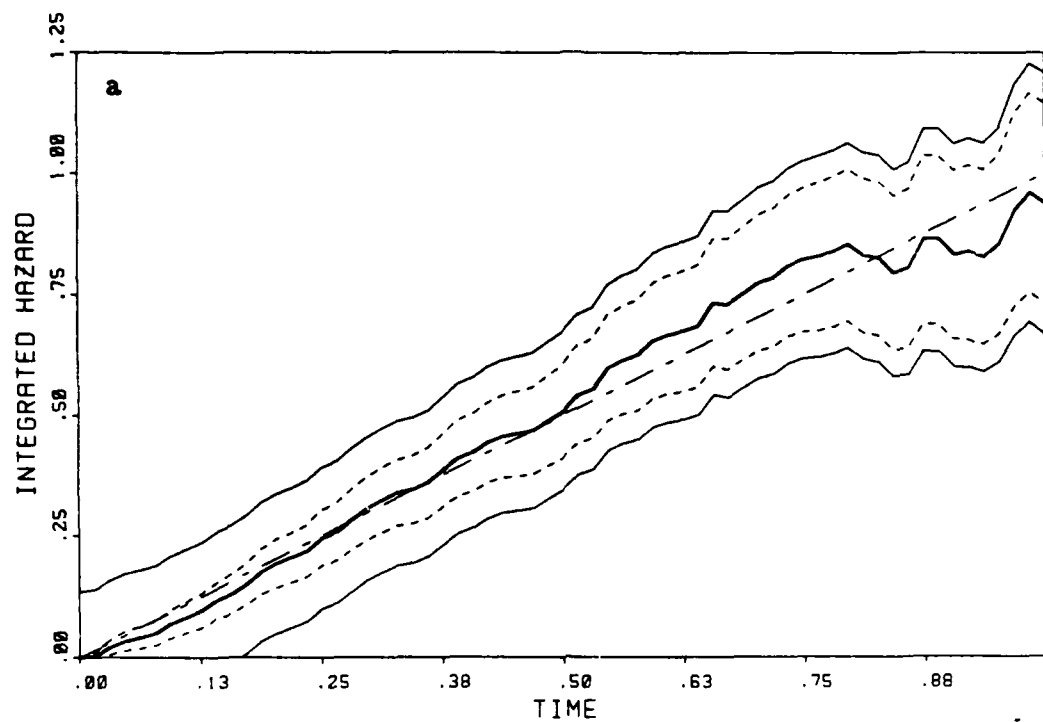


Figure 4. Heavy censoring (68%), $n = 2000$, $d = 64$. (a) Covariate 1. (b) Covariate 2.

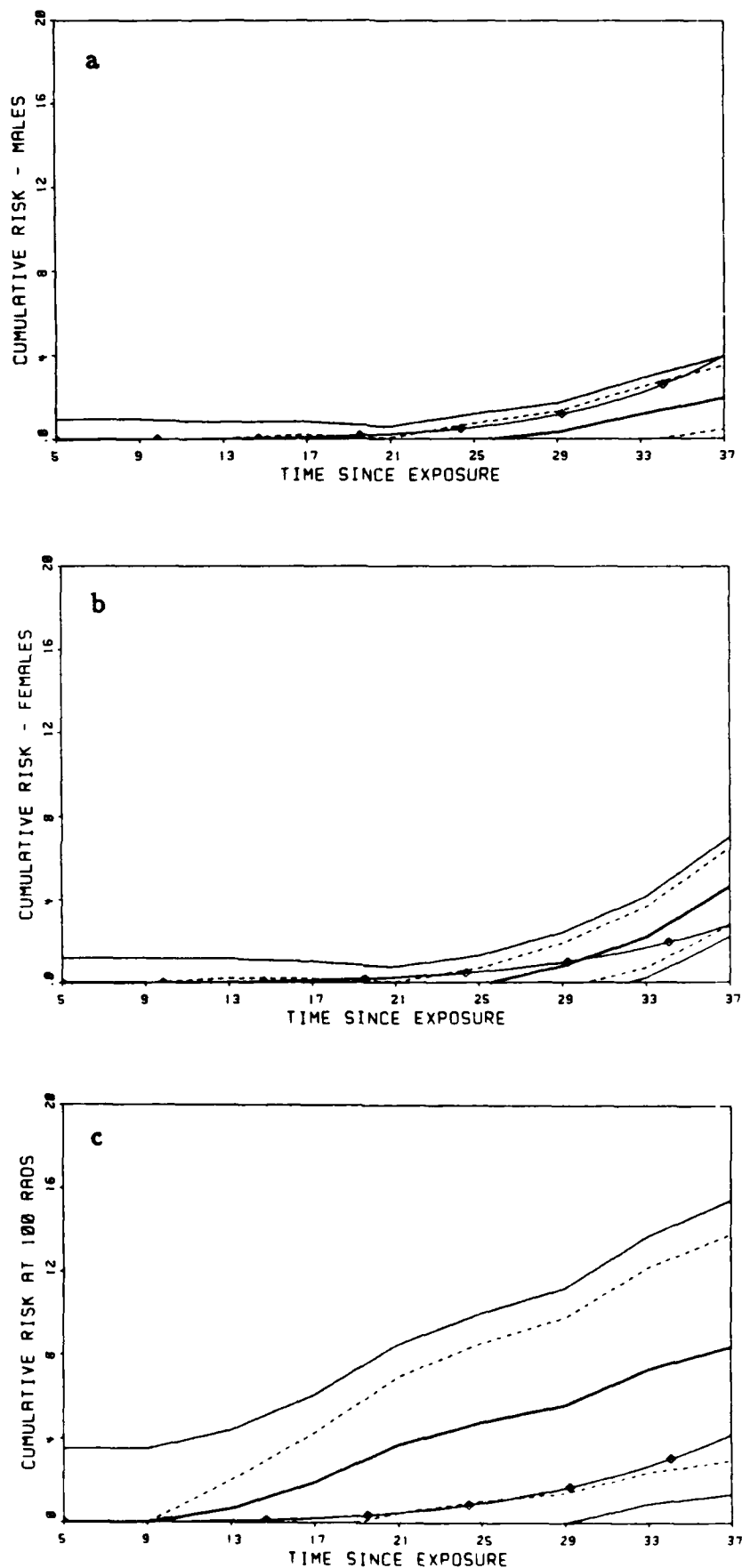


Figure 5. 0-9 Years of Age at Exposure. (a) Background for males. (b) Background for females. (c) Excess.

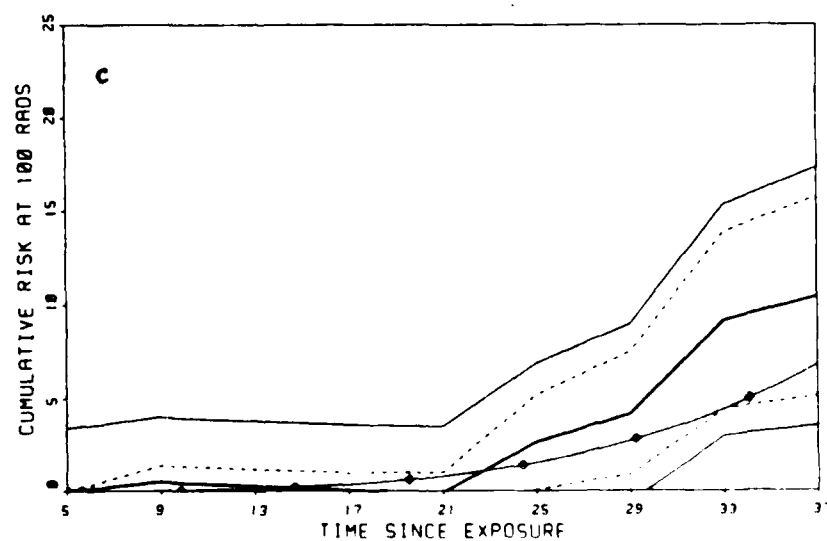
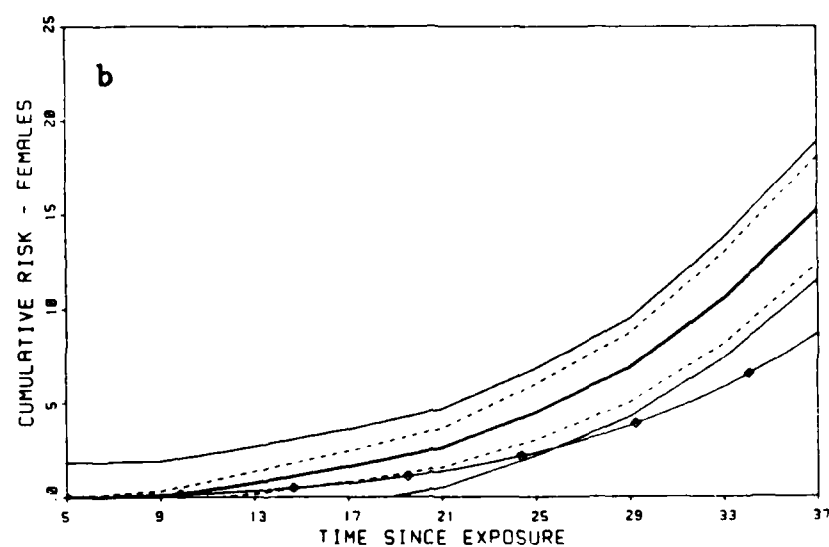
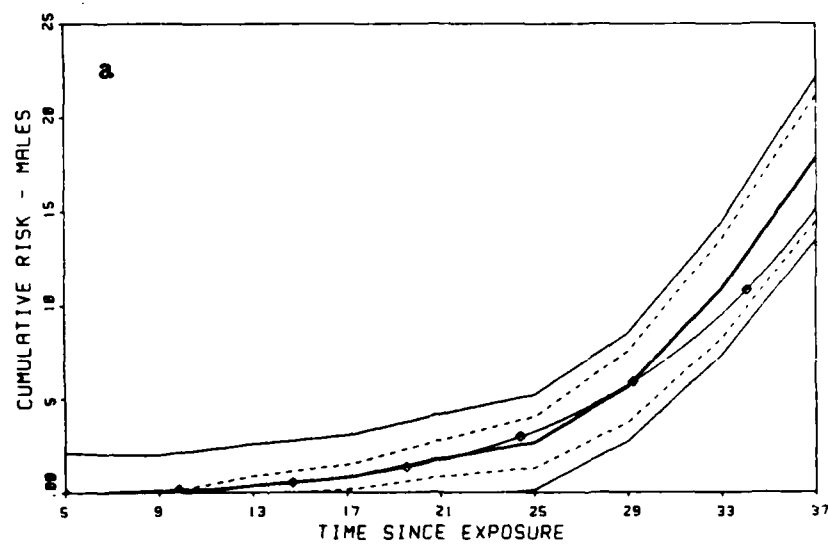


Figure 6. 10-19 Years of Age at Exposure. (a) Background for males. (b) Background for females. (c) Excess.

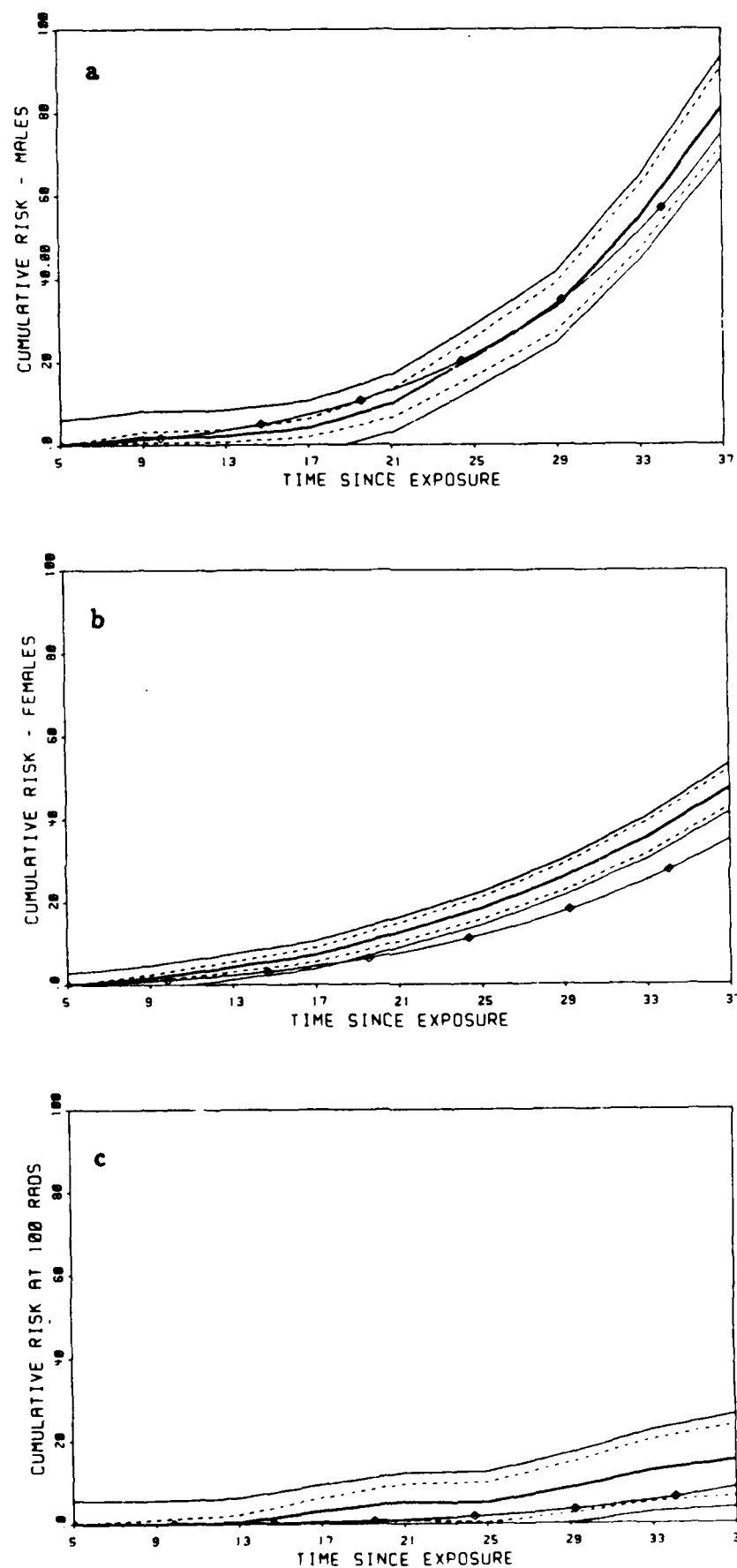


Figure 7. 20-34 Years of Age at Exposure. (a) Background for males. (b) Background for females. (c) Excess.

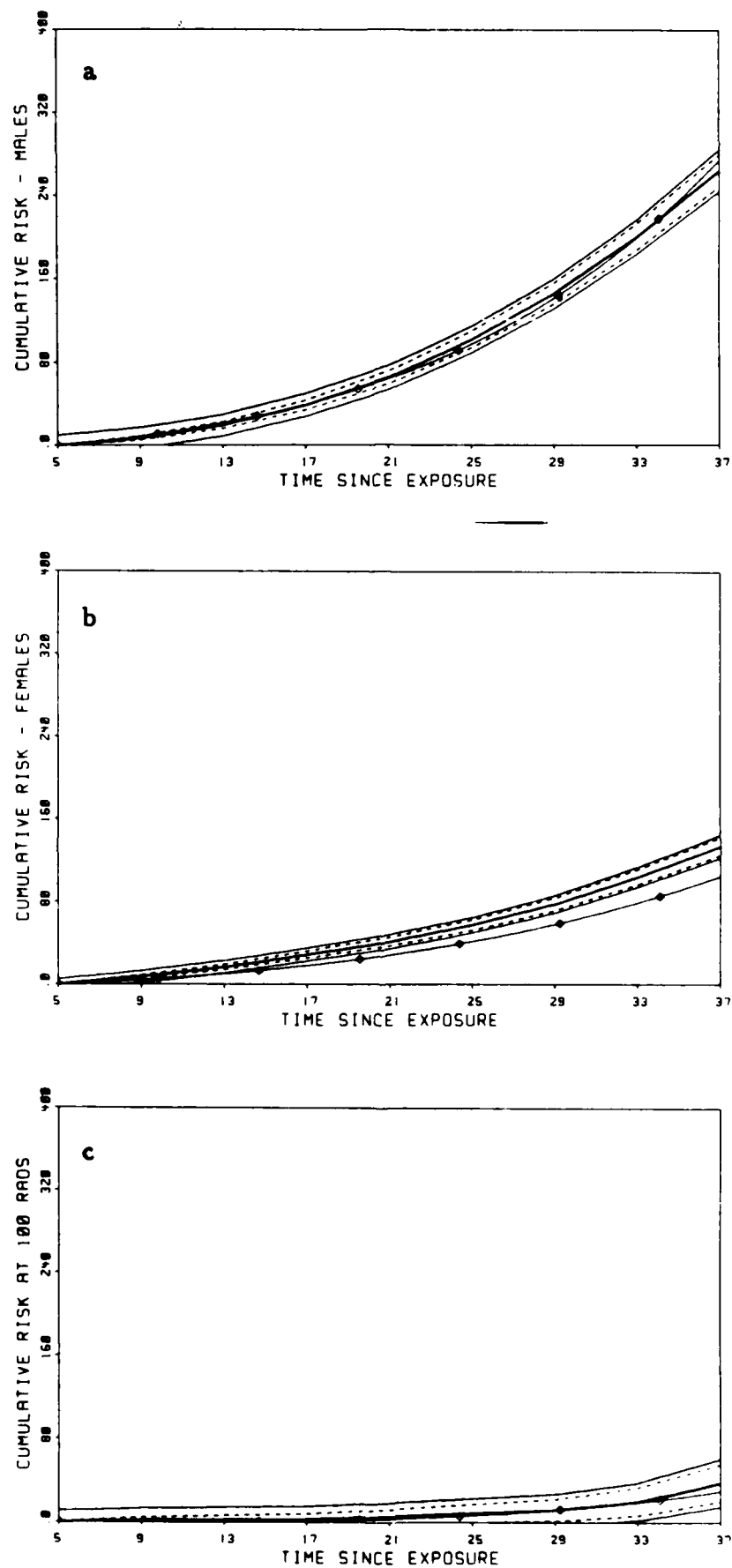


Figure 8. 35-49 Years of Age at Exposure. (a) Background for males. (b) Background for females. (c) Excess.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 396	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Survival Analysis Using Additive Risk Models		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Fred W. Huffer and Ian W. McKeague		8. CONTRACT OR GRANT NUMBER(s) N00014-86-K-0156
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 1111		12. REPORT DATE October 6, 1987
		13. NUMBER OF PAGES 36
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES Also issued as Army Research Office Technical Report No. D-96, Florida State University.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Grouped survival data, additive risk models, Monte-Carlo, atomic bomb survivors, nonproportional hazards, censoring, multivariate counting processes, martingale methods.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) PLEASE SEE FOLLOWING PAGE.		

20. ABSTRACT

Cox's (1972) proportional hazards model has so far been the most popular model for the regression analysis of censored survival data. However, the additive risk model of Aalen (1980) provides a useful and biologically more plausible alternative when large sample size makes its application feasible. Let $\lambda(t) = \lambda(t, Y)$ be the hazard function for a subject whose covariates are given by $Y = (Y_1, \dots, Y_p)'$. Aalen's model stipulates that $\lambda(t) = Y'\alpha(t)$, where $\alpha = (\alpha_1, \dots, \alpha_p)'$ is an unknown vector of hazard functions. This paper discusses inference for $\alpha_1, \dots, \alpha_p$ based on continuous and grouped data. Asymptotic distribution results are developed using the theory of counting processes and used to provide confidence bands for the cumulative hazard functions. The method is applied to data on the incidence of cancer mortality among Japanese atomic bomb survivors.

END
DATE
FILMED
MARCH
1988
DTIC